

MY WEIRD PROMPTS

Podcast Transcript

EPISODE #171

The Rise of AIO: Optimizing Your Website for AI Bots

Published January 05, 2026 • Runtime: 24:07

<https://myweirdprompts.com/episode/ai-optimization-llms-txt-future/>

EPISODE SYNOPSIS

In this episode of My Weird Prompts, Herman and Corn explore the radical shift from defensive SEO to the new frontier of AI Optimization (AIO). As we move into early 2026, the digital landscape has pivoted from building "digital fortresses" against scrapers to creating "machine-digestible" environments. The duo dives deep into the technical specifications of the llms.txt file—a markdown-based "cheat sheet" for large language models—and explains why providing a curated map of your site is essential for brand consistency. They discuss the evolving role of sitemaps and Schema.org as the "Rosetta Stone" for AI, ensuring that bots can cite your data with high confidence. From reducing latency for RAG systems to the concept of "responsive design for intelligence," this episode provides a comprehensive roadmap for anyone looking to make their content the primary source of truth in a conversational search world. Whether you are a developer, a marketer, or a researcher, understanding how to communicate directly with the neural networks is the ultimate inbound marketing strategy for the future.

DANIEL'S PROMPT

Daniel

While many tools allow website owners to block AI bots from scraping content, I believe there's a significant benefit to making websites as easy as possible for these bots to index in hopes of being cited as a source. What should website owners be doing to optimize their sites for AI indexing and citation? Specifically, what are the best practices for using tools like llms.txt, sitemaps, and metadata to leverage AI as an inbound marketing source?

TRANSCRIPT

Corn

Hey everyone, welcome back to My Weird Prompts. I am Corn, and I am sitting here in a very rainy Jerusalem morning with my brother.

Herman

Herman Poppleberry, present and accounted for. And yeah, the rain is really coming down today, Corn. It is the kind of weather that makes you want to just stay inside, drink far too much coffee, and dive into some deep technical rabbit holes.

Corn

Which is exactly what we are doing. Our housemate Daniel sent over a really interesting prompt this morning. He was listening to some of our recent discussions, like back in episode two seventy five when we talked about air gapped A I, and he started thinking about the opposite end of the spectrum. Instead of hiding data, how do we make it as attractive and digestible as possible for the bots?

Herman

It is a total shift in mindset, right? For the last couple of years, especially through twenty twenty four and twenty twenty five, the narrative was all about defensive posture. Everyone was terrified of the scrapocalypse. We saw all these tools from Cloudflare and other providers that were basically digital fortresses meant to keep the large language model crawlers out. But Daniel is pointing toward a new frontier. If you cannot beat the crawlers, why not make them your best friends?

Corn

Exactly. It is the evolution from S E O, search engine optimization, to what some people are calling A I O, or A I optimization. If the way people find information is shifting from a list of blue links to a conversational summary, then being the source for that summary is the new prime real estate.

Herman

It really is. And it is funny because we are recording this in early January of twenty twenty six, and the landscape has changed so much just in the last twelve months. We are seeing that the bots are not just mindless scrapers anymore. They are looking for high quality, structured context. They want to cite sources because the users are demanding it. If an A I gives an answer without a citation now, people tend to trust it less. So, the question is, how do you make your website the one that the A I trusts and chooses to cite?

Corn

That is the core of it. Daniel specifically mentioned a few tools like llms dot t x t, sitemaps, and metadata. I want to start with llms dot t x t because that feels like the newest and perhaps most misunderstood piece of this puzzle. Herman, you have been digging into the specifications for that, right?

Herman

Oh, I have been obsessed with it. It is such a simple, elegant solution to a massive problem. Think of llms dot t x t as the spiritual successor to robots dot t x t, but instead of telling bots where they cannot go, it provides a curated map of where they should go and what they will find there. It is essentially a markdown file that sits in the root directory of your website.

Corn

So, instead of a bot having to crawl every single page and try to figure out the hierarchy through trial and error, you are handing them a cheat sheet?

Herman

Precisely. The specification for llms dot t x t is designed to be extremely lightweight. It uses markdown because large language models are incredibly good at parsing markdown. It is their native tongue, in a way. In that file, you provide a high level summary of the site, and then you provide links to the most important pages, often with a brief description of what each page contains. But here is the kicker, you can also point to an llms hyphen full dot t x t file.

Corn

And what is the difference there? Is that just the long form version?

Herman

Yeah, exactly. The full version can actually contain the full text of your most important documentation or articles, all concatenated into one single, easily digestible file. Imagine you are a developer or a researcher. Instead of having to click through twenty different pages to understand a concept, the A I can just ingest that one full text file in a single request. It reduces the latency for the bot, reduces the load on your server, and significantly increases the chances that the A I will have the full context it needs to give an accurate answer.

Corn

That is fascinating because it turns the traditional web design philosophy on its head. We usually design for humans, with lots of whitespace, images, and nested menus. But an A I does not care about your beautiful C S S or your hero image. It wants the raw, structured data. Does this mean we are moving toward a world where every site has a shadow version of itself just for the machines?

Herman

In a way, yes. But it is a collaborative shadow. If you look at how the big players are using this in twenty twenty six, it is not just about the text. It is about the intent. When you format your llms dot t x t file, you are essentially saying, here is the objective truth about my product or my research. By providing that structured path, you are preventing the A I from hallucinating based on outdated information it might have found in a cache somewhere from three years ago.

Corn

I can see how that would be huge for brand consistency. If I am a company and I just released a new version of my software, I want the A I to know about the new features immediately. But I am curious about the citation aspect. How does making it easy for the bot lead to that inbound marketing win Daniel was talking about?

Herman

That comes down to how these models handle their retrieval augmented generation, or R A G. When an A I gets a query, it looks through its indexed data for the most relevant snippets. If your data is cleanly formatted in a way that perfectly matches the query, the model is more likely to pull your snippet into its context window. And because you provided the clear structure and the direct links in your llms dot t x t file, the model has a clear path to attribute that information back to you. It is like providing a bibliography for the A I before it even writes the essay.

Corn

It makes total sense. It is about reducing friction. The less work the A I has to do to verify your information, the more likely it is to use it. But before we get too deep into the metadata and sitemaps, we should probably take a quick break for our sponsors.

Herman

Good idea. I need a quick refill on my tea anyway.

Corn

We will be right back. Larry: Are you tired of your website being ignored by the digital elite? Does your server feel lonely and unloved? Introducing the Algorithm Whisperer Pendant. This handcrafted, copper plated charm is infused with the essence of high frequency trading data and the tears of a thousand failed startups. Simply hang the pendant over your router, and watch as your search rankings defy the laws of physics. Our patented resonance technology speaks directly to the hidden layers of the neural networks, convincing them that your blog about artisanal pickles is actually the foundational text of a new civilization. Side effects may include spontaneous coding in ancient Greek, a sudden fear of magnets, and the ability to hear the sound of fiber optic cables. The Algorithm Whisperer Pendant. Because why optimize your content when you can just haunt the machine? BUY NOW!

Corn

Alright, thanks Larry. I am not sure hanging a copper pendant over a router is exactly what Daniel had in mind, but hey, to each their own.

Herman

I love how Larry always manages to find a way to make technology sound like nineteenth century occultism. It is a gift, really.

Corn

It really is. Anyway, back to the world of actual optimization. We talked about llms dot t x t, but Daniel also mentioned sitemaps and metadata. Now, sitemaps have been around since the beginning of time in internet years. Why are they suddenly important for A I?

Herman

Well, the reason they are important now is that A I crawlers are using them differently than traditional search engines. Google or Bing might use a sitemap to discover new pages, but they still rely heavily on the link graph, following one page to the next. A I models, especially the ones running real time web searches like we see in twenty twenty six, are often looking for the freshest data possible. A well structured S M L sitemap with clear last modified timestamps tells the bot exactly what has changed since its last training run.

Corn

So it is about the delta. It is about showing the bot what is new so it does not waste resources on the old stuff.

Herman

Exactly. And you can actually extend your sitemap with custom tags. We are starting to see people use specialized schema within their sitemaps to indicate which pages are high density information pages versus marketing landing pages. If I am a bot, I want to prioritize the documentation and the white papers over the sign up page. By flagging those in the sitemap, you are guiding the A I toward the content that is most likely to be cited as a source of truth.

Corn

That leads perfectly into the metadata discussion. We have been using things like Open Graph for social media and Schema dot org for rich snippets in search results for years. How does that translate to the A I era?

Herman

It translates because Schema dot org is basically the Rosetta Stone for A I. When you use JSON L D to mark up your content, you are giving the A I a set of key value pairs that it can understand with zero ambiguity. If you have an article about a specific chemical compound, and you use the appropriate Schema markup to define the compound's properties, the A I does not have to guess if that number is a boiling point or a molecular weight. It knows.

Corn

And that certainty is what drives citations.

Herman

Precisely. AI models are built on probabilities. If the model is ninety nine percent sure that your site has the correct answer because it is clearly labeled in the metadata, it is going to pick you over a site where it only has seventy percent confidence in its extraction. This is especially true for things like prices, dates, and technical specifications. If you want to be the source of truth, you have to label your truth clearly.

Corn

I think there is a second order effect here that is really interesting for marketing. If you are optimized for these bots, you are not just getting a link in a chat box. You are potentially becoming part of the AI's internal knowledge base during its next training cycle.

Herman

That is a great point, Corn. We are seeing a shift where the line between search and training is blurring. In twenty twenty six, many models are doing continuous or incremental training. If your site is easy to index, your brand's perspective and your data become part of the model's fundamental understanding of a topic. That is the ultimate inbound marketing. You are not just being found, you are being remembered.

Corn

It is like the difference between being a book on a shelf that someone might pick up, and being part of the person's actual education. But let's talk about the practical side for a second. If I am a website owner, and I want to implement this today, what does my checklist look like?

Herman

First step, without a doubt, is creating that llms dot t x t file. It is the lowest hanging fruit. You can manually curate it in about twenty minutes. List your core pages, give them clear, descriptive titles, and provide a summary of what your site is about. Use natural language but keep it concise. Remember, you are talking to a machine that likes clarity.

Corn

And what about the markdown aspect? I know you mentioned that earlier.

Herman

Yeah, this is a big one. Ensure that the pages you are linking to are also easy to parse. Avoid complex layouts with lots of nested divs if you can. If you can provide a text only or markdown version of your long form content, do it. Some people are even using a media query or a specific header to serve a simplified version of the page when they detect an A I crawler. It is like a responsive design, but for intelligence instead of screen size.

Corn

That is a brilliant analogy. Responsive design for intelligence. I like that. What about the citation bait? How do you actually structure a sentence or a paragraph to make it more quotable for an A I?

Herman

I call this the Verifiable Factoid method. A I models love sentences that follow a clear subject, predicate, object structure with specific data points. Instead of saying, our software is really fast and helps you work better, say, our software reduces latency by forty five percent compared to industry standards as of January twenty twenty six. The second sentence is a fact that can be extracted, stored, and cited. The first sentence is just fluff.

Corn

So, it is about being more like an encyclopedia and less like a brochure.

Herman

Exactly. The brochure is for the human who wants to feel an emotion. The encyclopedia is for the A I that wants to provide an answer. To win at A I O, you have to be both. You have to have the human centric landing page, but you need to have that data rich foundation that the bot can grab onto.

Corn

It is also worth mentioning that this is not just for tech companies. If you are a local restaurant in Jerusalem, you should have your menu marked up with Schema. You should have an llms dot t x t file that clearly states your hours, your location, and your signature dishes. When someone asks an A I, where can I find the best hummus in Jerusalem that is open on a Tuesday morning, you want the A I to have zero doubt that you are the right answer.

Herman

Right. And think about the implications for trust. In episode one hundred eleven, we talked about benchmarking and the word error rate in A S R tools. Reliability was the key theme there. It is the same thing here. If an A I cites you and the user clicks through and finds that the information is accurate and easy to find, that builds a massive amount of trust not just with the user, but with the model's feedback loop. These models are constantly being fine tuned based on user satisfaction. If your site consistently leads to satisfied users, you become a preferred source.

Corn

It is a virtuous cycle. You make it easy for the bot, the bot gives a good answer, the user is happy, the bot's developers see that the source was high quality, and your ranking in the latent space of that model goes up.

Herman

I love that term, ranking in the latent space. That is exactly what it is. We are moving away from a world of keywords and toward a world of semantic vectors. Your goal as a website owner is to position your brand as close as possible to the concepts you want to be associated with.

Corn

So, we have talked about the technicals and the strategy. Let's look at some of the misconceptions. I think a lot of people still think that if they let the bots in, they are just giving away their value for free. How do you counter that argument in twenty twenty six?

Herman

It is a valid concern, but it is also a bit of a dinosaur mentality. If your value is just the raw information, then yeah, you might be in trouble. But for most businesses, the value is in the service, the expertise, or the community. By letting the bot index your information, you are using that information as a lighthouse to bring people to your actual value. If you block the bot, you are not protecting your information, you are just making yourself invisible. In a world where eighty percent of information seeking starts with an A I prompt, being invisible is the same as being out of business.

Corn

It is the difference between a walled garden and a storefront. A walled garden is great if you already have a loyal following, but if you want new customers, you need a storefront that people can actually see from the street. And in this case, the street is the A I's response window.

Herman

That is a perfect way to put it. And let's not forget about the legal landscape. We are seeing more and more cases where A I companies are willing to pay for high quality data, but they are only going to pay for data that is easy to ingest and verify. By following these best practices, you are essentially making your data enterprise ready. Whether you are looking for citations or a licensing deal, the requirements are the same: structure, clarity, and accessibility.

Corn

It is funny how we have come full circle. In the early days of the web, it was all about being open and connected. Then we got into this era of silos and paywalls and defensive S E O. Now, thanks to A I, we are being forced to go back to those original principles of clear, structured communication.

Herman

It is a return to the semantic web that Tim Berners Lee envisioned decades ago. We just needed a machine smart enough to actually use it.

Corn

Well, I think we have given Daniel a lot to chew on. To recap: get that lms dot t x t file up, use Schema dot org metadata for everything, keep your sitemaps fresh, and start writing in a way that is as easy for a bot to cite as it is for a human to read.

Herman

And don't forget to keep an eye on the emerging standards. This field is moving so fast. What we are talking about in January might be supplemented by new protocols by June. Stay curious and stay flexible.

Corn

Absolutely. And hey, if you are listening to this and you found it helpful, we would really appreciate it if you could leave us a review on Spotify or whatever podcast app you use. It genuinely helps other people find the show, and we love hearing from you.

Herman

Yeah, it really does make a difference. We have been doing this for two hundred seventy eight episodes now, and the community feedback is what keeps us going.

Corn

For sure. You can also find us at my weird prompts dot com. We have got the R S S feed there and a contact form if you want to send in a prompt like Daniel did. We are always looking for new rabbit holes to explore.

Herman

Just maybe don't send us any prompts about copper pendants. I think Larry has that market covered.

Corn

Fair enough. Well, this has been My Weird Prompts. I am Corn.

Herman

And I am Herman Poppleberry.

Corn

Thanks for listening, and we will talk to you next week.

Herman

See ya!